

# **ANALISIS BUTIR SOAL UJIAN AKHIR SEMESTER GANJIL (UAS) BAHASA JERMAN KELAS X SMAN 12 SURABAYA**

**Elsa Wandira Maulidatirrahma**

Mahasiswa Program Studi S1 Pendidikan Bahasa Jerman, Fakultas Bahasa dan Seni,  
Universitas Negeri Surabaya

[elsamaulidatirrahma16020094033@mhs.unesa.ac.id](mailto:elsamaulidatirrahma16020094033@mhs.unesa.ac.id)

**Drs. Suwarno Imam Samsul, M.Pd.**

Dosen Program Studi S1 Pendidikan Bahasa Jerman, Fakultas Bahasa dan Seni, Universitas  
Negeri Surabaya

## **ABSTRAK**

Untuk mengukur sejauh mana kemampuan dan pemahaman peserta didik terhadap materi yang telah diberikan selama 1 semester, maka di akhir semester pendidik akan memberikan sebuah tes kepada peserta didik yang kemudian dari hasil tersebut dapat memberikan sebuah gambaran terhadap pendidik seberapa jauh keberhasilan capaian tujuan pembelajaran. Sebuah tes dapat dikatakan baik atau layak apabila tes tersebut dapat dipertanggung jawabkan kesahihan, ketafsiran, kelayakan, kebergunaan, keterpercayaan, maupun efektivitas butir soal yang meliputi tingkat kesulitan dan daya beda yang baik, karena tidak semua butir soal dapat sesuai dengan kriteria tersebut. Rumusan masalah dalam penelitian ini adalah bagaimanakah tingkat kesulitan butir soal, daya beda dan efektivitas distraktor jawaban dari soal Ujian Akhir Semester Ganjil (UAS) bahasa Jerman kelas X SMAN 12 Surabaya TP 2019/2020. Penelitian yang saya lakukan di SMAN 12 Surabaya ini merupakan penelitian kuantitatif dengan cara menganalisis tingkat kesulitan, daya beda dan efektivitas distraktor dengan menggunakan rumus *IF (Indeks Facility)* dan *ID (Indeks Discrimination)*. Dari hasil penelitian yang telah dilakukan, diketahui bahwa dari 50 butir soal pilihan ganda Ujian Akhir Semester Ganjil bahasa Jerman kelas X SMAN 12 Surabaya terdapat 21 butir soal telah memenuhi syarat ITK, dan 13 butir soal telah memenuhi syarat IDB. Dan hasil akhir dari perhitungan indeks tingkat kesulitan dan indeks daya beda diketahui hanya 12 butir soal yang dapat dinyatakan layak. Dari hasil analisis tingkat kesulitan butir soal, telah terbagi 3 golongan soal yaitu mudah, sedang dan sulit. Dari UAS ini diketahui 3 soal tergolong sulit, 7 soal tergolong sedang dan 40 soal tergolong mudah. Namun dari 40 soal tergolong mudah ini telah di analisis bahwa 29 soal tergolong terlalu mudah sehingga dinyatakan tidak berfungsi. Untuk hasil analisis Indeks Daya Beda telah diketahui terdapat 37 butir soal yang tidak layak dan diketahui juga dari soal tidak layak tersebut terdapat 13 butir soal yang indeks daya beda yang minus yaitu kelompok rendah lebih banyak menjawab benar daripada kelompok tinggi sehingga 13 soal tersebut tidak bisa di revisi dan diharapkan untuk dibuang dan diganti soal yang layak. Untuk sisa soal yang dinyatakan tidak layak dapat diperbaiki atau di revisi. Hasil Analisis Distraktor dari 250 opsi jawaban dan pengecoh, menunjukkan 133 opsi jawaban yang berfungsi baik, dan 117 tidak berfungsi baik.

*Kata kunci : analisis butir soal, tingkat kesulitan, daya beda dan distraktor.*

## ABSTRACT

To measure the extent to which the ability and understanding of students of the material that has been given during 1 semester, then at the end of the semester the educator will give a test to students who then from the results can provide an overview of the educator how far the success of learning objectives. A test can be said to be good or feasible if the test can be accounted for validity, interpretation, appropriateness, usefulness, trustworthiness, and effectiveness of items that include a good level of difficulty and different power, because not all items can be in accordance with these criteria. The formulation of the problem in this research is how the level of difficulty of the items, the different power and effectiveness of the answer distractor from the German Final Examination (UAS) class X in SMAN 12 Surabaya TP 2019/2020. The research that I did at SMAN 12 Surabaya is a quantitative study by analyzing the level of difficulty, different power and effectiveness of the distractor by using the formula IF (Index Facility) and ID (Index Discrimination). From the results of the research that has been carried out, it is known that of the 50 multiple choice questions at the end of the German language semester examination at SMAN 12 Surabaya there are 21 items that have met ITK requirements, and 13 items have met IDB requirements. And the final results of the calculation of the difficulty level index and the different power index are known only 12 items that can be declared feasible. From the results of the analysis of the difficulty level of the items, 3 groups of questions were divided namely easy, medium and difficult. From this UAS, 3 problems are classified as difficult, 7 questions are classified as moderate and 40 questions are classified as easy. However, of the 40 questions classified as easy, it has been analyzed that 29 questions were classified as too easy to declare not functioning. Then for the results of the analysis of the different power index it is known that 37 items are not feasible, and it is also known from the improper questions that there are 13 items with a minus power index difference, namely the low group answers more correctly than the high group so that 13 questions cannot be revised and expected to be discarded and replaced with appropriate questions. For some of the remaining questions, it can be revised or corrected so that it is feasible. Whereas for the analysis of the distractors provided in these 50 items, there were 250 answer and deception options, and there were 133 answer options that functioned well, and the remaining 117 did not function properly because none of the test participants chose any of the answer options. For answer options that don't work well, you can fix them.

*Keywords: item analysis, level of difficulty, different power and distracto*

## PENDAHULUAN

Pada umumnya untuk mengukur tingkat kemampuan dan pemahaman peserta didik, pendidik memberikan tes kepada peserta didik kemudian dari hasil tes tersebut dapat didiagnosis dimana letak kesulitan belajar peserta didik serta dapat mengetahui kemampuan pengetahuan peserta didik, dan memberikan informasi tentang keberhasilan pencapaian tujuan pembelajaran peserta didik. Tetapi untuk dapat mengukur tersebut tes harus dikatakan layak atau baik. Ciri tes yang baik haruslah dipertanggung jawabkan mulai dari kesahihan (validitas), kelayakan, ketertafsiran, keterpercayaan (reliabilitas), kebergunaan maupun efektivitas butir soal meliputi tingkat kesulitan dari soal tersebut serta daya pembeda soal yang baik.

SMAN 12 Surabaya memberikan peserta didik mereka tes Ujian Akhir Semester Ganjil (UAS) untuk mengetahui dan mengukur Ganjil sejauh mana kemampuan peserta didik SMAN 12 Surabaya menguasai materi dalam satu semester. Maka dari itu kualitas soal juga perlu diuji, yaitu dengan uji terhadap tingkat kesukaran, daya pembeda, dan fungsi distraktor item butir soal yang akan diujikan. Penelitian ini dilakukan di SMAN 12 Surabaya karena di sekolah ini belum pernah dilakukan analisis butir soal terhadap soal Ujian Akhir Semester Ganjil.

Rumusan masalah pada penelitian ini adalah : bagaimanakah Indeks Tingkat Kesulitan, Indeks Daya Beda dan Efektivitas Distraktor pada tiap butir soal. Penelitian ini bertujuan untuk mendeskripsikan Indeks Tingkat Kesulitan, Indeks Daya Beda dan Efektivitas Distraktor pada tiap butir soal Ujian Akhir Semester Ganjil SMA Negeri 12 Surabaya. Batasan masalah dalam penelitian ini adalah penelitian ini difokuskan untuk mengetahui Indeks Tingkat Kesulitan, Indeks Daya Beda dan Efektivitas Distraktor pada tiap butir soal multiple choice dengan menggunakan analisis kuantitatif. Objek dalam penelitian ini difokuskan untuk soal UAS bahasa Jerman kelas X SMAN 12 Surabaya.

## METODE

Jenis penelitian yang dilakukan di SMAN 12 Surabaya adalah jenis penelitian kuantitatif. Seperti yang dikatakan Arikunto (2002:10) penelitian kuantitatif adalah sebuah penelitian yang hasil pengukurannya dituntut banyak menggunakan angka, seperti proses pengumpulan data hingga penampilan dari hasilnya. Metode penelitian ini dilakukan secara analisis kuantitatif dengan prosedur penelitian, yaitu:

1. Tahap persiapan

Menggunakan studi pustaka yaitu mencari dan menumpulkan teori yang sesuai dengan pokok permasalahan yang diteliti dan membuat izin pengambilan data.

2. Tahap pengumpulan data

Data diambil di SMAN 12 Surabaya kelas X tahun pelajaran 2019/2020. Data tersebut berupa lembar soal serta kunci jawaban.

3. Tahap pengolahan data

Data yang terkumpul kemudian dianalisis dengan menggunakan teknik analisis statistik yang bersifat kuantitatif dan kemudian dideskripsikan untuk revisi.

4. Tahap penyusunan laporan

Setelah semua data di analisis maka dilakukan penyusunan laporan untuk menjawab tujuan penelitian.

Analisis Kuantitatif dalam penelitian ini adalah untuk menganalisis tingkat kesulitan dan daya pembeda soal dengan rumus.

Langkah-langkah yang dilakukan untuk menganalisis butir soal menurut Nurgiyantoro (2001:136) adalah sebagai berikut :

- a. Mengurutkan skor pada lembar jawaban siswa dari skor yang tertinggi berturut-turut sampai yang terbawah.
- b. Setelah data sudah urut, lalu dibagi menjadi 2 kelompok yaitu kelompok tinggi dan kelompok rendah. Misal jumlah murid adalah 40 orang, maka 20 siswa dengan nilai tertinggi adalah kelompok tinggi, dan 20 murid dengan nilai terendah adalah kelompok rendah.
- c. Setelah 2 kelompok sudah terbentuk, yaitu kelompok tinggi dan kelompok

rendah selanjutnya bisa menghitung Indeks Tingkat Kesulitan (ITK) butir soal yang bisa dilakukan dengan melihat tabel jawaban peserta didik, dengan rumus berikut:

$$ITK = \frac{FKT + FKR}{N}$$

Keterangan :

ITK = Indeks Tingkat Kesulitan yang dicari

FKT = Jumlah jawaban benar kelompok tinggi

FKR = Jumlah jawaban benar kelompok rendah

N = Jumlah peserta tes kedua kelompok.

Setelah selesai menghitung Indeks Tingkat Kesulitan (ITK), dapat dilanjutkan dengan menghitung Indeks Daya Beda (IDB) dengan rumus seperti berikut :

$$IDB = \frac{FKT - FKR}{N}$$

Keterangan :

IDB = Indeks Daya Beda yang dicari

FKT = Jumlah jawaban benar kelompok tinggi

FKR = Jumlah jawaban benar kelompok rendah

N = Jumlah subjek kelompok tinggi atau rendah

Setelah menemukan Indeks Tingkat Kesulitan (ITK) dan Indeks Daya Beda (IDB) butir soal, maka langkah selanjutnya ialah mencari Efektifitas Distraktor yaitu dilakukan dengan cara membuat tabel penyebaran frekuensi jawaban peserta didik kelompok tinggi dan kelompok rendah. Berikut adalah contoh tabel nya :

butir soal	Kelompok Tinggi			
	A	B	C	D

1	*10	2	2	2
2	5	-	-	*15
3	6	*7	3	1
4	7	*6	6	2
5	9	-	*7	5
Dst...				

butir soal	Kelompok Rendah			
	A	B	C	D
1	*10	4	6	2
2	8	*7	3	3
3	7	2	4	*6
4	6	*8	2	4
5	3	5	*7	5
Dst...				

Catatan :

- A,B,C,D = Opsi jawaban yang telah disediakan

- (\*) = jawaban yang betul

- Diatas adalah contoh 20 murid kelompok tinggi dan 20 murid kelompok rendah.

Setelah data-data sebaran pilihan jawaban oleh peserta didik telah didapatkan, maka selanjutnya kita akan dapat melakukan analisis distraktor tiap butir soal untuk mengetahui efektif atau tidaknya distraktor atau pengecoh tersebut. Menurut Nurgiyantoro (2010:2010) ada beberapa kriteria distraktor yang efektif, yaitu :

1. Semua distraktor (opsi salah) harus ada yang memilih
2. Jumlah memilih opsi-salah dari peserta kelompok tinggi harus lebih sedikit dari kelompok rendah
3. Jika memilih opsi-salah hanya satu, ia harus dari kelompok rendah
4. Kriteria 2 dan 3 terkadang dipandang memberatkan, yang pada intinya tidak berbeda dengan logika tuntutan IDB diatas, maka yang digunakan secara efektif hanyalah kriteria 1



Opsi-opsi distraktor yang kurang efektif harus dibuang atau direvisi. Opsi-opsi distraktor yang efektif adalah opsi yang tidak mencolok dan hampir mirip dengan jawaban aslinya, atau “serupa tapi tak sama” sehingga peserta didik yang kurang teliti akan memilih opsi distraktor tersebut.

## PEMBAHASAN

Analisis butir soal memiliki tujuan yakni, untuk mengidentifikasi tiap soal dari soal-soal yang baik, kurang, maupun soal yang jelek. Untuk mengetahui kualitas soal tersebut maka harus diketahui Indeks tingkat Kesulitan (ITK) dan Indeks Daya Beda (IDB). Untuk mengetahui ITK dan IDB butir soal harus dilakukan dengan beberapa langkah, yaitu :

1. Langkah pertama yang dilakukan adalah membuat tabel yang berisi inisial subyek peserta didik serta berisi nilai dan jawaban benar atau salah dari hasil koreksi yang didapatkan oleh masing-masing peserta didik secara keseluruhan dengan memberi skor 1 untuk jawaban benar dan skor 0 untuk jawaban yang salah.
2. Langkah kedua yaitu membuat tabel yang berisi inisial subyek peserta didik lalu diurutkan dari peserta didik yang memiliki jumlah jawaban benar paling banyak atau jumlah kesalahan paling sedikit hingga peserta didik yang memiliki jumlah jawaban benar paling sedikit atau memiliki jumlah kesalahan paling banyak.
3. Setelah data hasil nilai peserta didik sudah diurutkan mulai dari yang paling banyak memiliki jawaban paling benar atau memiliki kesalahan sedikit hingga ke jawaban benar yang paling sedikit atau lebih banyak jawaban salahnya, lalu setelah itu bisa dibagi menjadi 2 kelompok. Yaitu kelompok tinggi dan kelompok rendah. Kelompok tinggi yaitu setengah dari jumlah keseluruhan peserta didik yang memiliki nilai

paling tinggi, sementara kelompok rendah adalah sisa dari setengah kelompok tinggi. Jumlah peserta didik dalam satu kelas terdiri dari 36 siswa, sehingga urutan siswa nomor 1-18 teratas adalah kelompok tinggi, sedangkan siswa nomor urut 19-36 adalah kelompok rendah.

4. Setelah membentuk kelompok tinggi dan kelompok rendah, maka langkah selanjutnya yang akan dikerjakan adalah mengetahui Indeks Tingkat Kesulitan (ITK) dengan menghitungnya menggunakan rumus.
5. Setelah menghitung dan mengetahui indeks Tingkat Kesulitan butir soal, maka selanjutnya bisa dilanjutkan dengan menghitung indeks Daya Beda (IDB) butir soal menggunakan rumus.
6. Setelah menghitung Indeks Tingkat Kesulitan (ITK) dan Indeks Daya Beda (IDB) selanjutnya adalah membuat tabel gabungan hasil perhitungan ITK dan IDB, dari tabel tersebut dapat dilihat soal yang layak dan tidak. Soal yang layak adalah soal yang dapat memenuhi syarat ITK dan IDB, jika hanya dapat memenuhi syarat ITK dan tidak dapat memenuhi syarat IDB maka soal tersebut dikatakan tidak layak, begitu pula sebaliknya. Untuk dapat memenuhi syarat ITK, tiap butir soal harus memiliki ITK 0,25-0,85 semakin tinggi ITK maka semakin mudah soal tersebut, semakin rendah ITK maka semakin sulit soal tersebut. Sedangkan untuk dapat memenuhi syarat IDB, tiap butir soal harus memiliki IDB diatas 0,25 semakin tinggi IDB maka soal tersebut semakin berfungsi baik karena dapat membedakan kelompok tinggi dan rendah. Untuk soal yang tidak dapat memenuhi syarat ITK atau IDB dapat direvisi atau diperbaiki.

## HASIL

Setiap butir soal dinyatakan layak ataupun tidak dapat dilihat dari ITK (Indeks Tingkat

Kesulitan) dan IDB (Indeks Daya Beda) yang telah atau belum memenuhi syarat yang ditentukan. Apabila suatu butir soal telah memenuhi syarat Indeks Tingkat kesulitan namun tidak dapat memenuhi syarat Indeks Daya Beda maka butir soal tersebut dianggap tidak berfungsi atau tidak layak. Begitupula sebaliknya, jika butir soal tersebut dapat memenuhi syarat Indeks Daya Beda namun tidak dapat memenuhi syarat Indeks Tingkat Kesulitan maka butir soal tersebut pun dianggap tidak berfungsi atau tidak layak. Jadi setiap butir soal yang dianggap layak dan baik yaitu yang memiliki ITK dan IDB yang telah memenuhi syarat. Sehubungan dengan itu maka hasil analisis yang telah dilakukan untuk mencari Indeks Tingkat Kesulitan, Indeks Daya Beda dan analisis Distraktor, maka ditemukan beberapa hasil yakni ada sebagian butir soal yang dapat dikatakan layak atau berfungsi dan ada juga butir soal yang dapat dikatakan tidak berfungsi dengan baik. Maka selanjutnya yang dapat dilakukan adalah me revisi atau mengganti soal soal yang dianggap tidak layak dan tidak berfungsi atau butir soal yang ITK dan IDB nya belum memenuhi syarat. Sedangkan untuk butir soal yang telah memenuhi syarat ITK dan IDB nya tidak perlu di revisi atau diganti lagi.

Sesuai dengan hasil analisis yang telah dilakukan, maka telah diketahui dari total 50 butir soal , terdapat 29 butir soal tidak memenuhi syarat Indeks Tingkat Kesulitan, kemudian diketahui juga 37 butir soal tidak memenuhi syarat Indeks Daya Beda. Hasil akhir seluruhnya maka telah diketahui ada 12 butir soal yang telah memenuhi syarat ITK (Indeks Tingkat Kesulitan) dan IDB (Indeks Daya Beda) ,maka 12 butir soal tersebut dapat dikatakan layak, dan sisa 38 butir soal yang dianggap tidak layak karena tidak dapat memenuhi syarat ITK dan IDB , maka 38 butir soal tersebut dapat di revisi atau diganti dengan soal yang lebih baik lagi. Ada 29 butir soal yang terdapat di soal ujian Akhir

Semester Ganjil SMAN 12 Surabaya yang tidak memenuhi syarat Indeks Tingkat Kesulitan, maka dapat disimpulkan bahwa soal ujian Akhir Semester Ganjil ini masih terlalu mudah bagi peserta didik. Kemudian terdapat 37 butir soal yang tidak dapat memenuhi syarat Indeks Daya beda, maka dapat disimpulkan bahwa soal Ujian Akhir Semester Ganjil ini kurang berfungsi baik untuk dapat membedakan peserta didik kelompok tinggi dan peserta didik kelompok rendah.

Terdapat beberapa butir soal yang hasil analisis ITK dan IDB nya hampir mencapai kriteria persyaratan, maka butir soal tersebut masih bisa direvisi kembali. Namun terdapat juga butir soal yang memang benar-benar harus diganti bahkan dibuang karena perolehan ITK dan IDB nya jauh dari layak. Berikut adalah beberapa butir soal yang perlu di revisi atau dibuang :

- Butir soal yang terlalu mudah dengan perolehan ITK 0,85-1,00 :  
1, 3, 4, 5, 6, 9, 11, 12, 13, 14, 15, 16, 17, 21, 22, 23, 24, 27, 34, 35, 36, 39, 40, 41, 42, 44, 47, 49, 50.
- Butir soal dengan perolehan IDB kurang dari 0,25 :  
1, 3, 4, 5, 6, 7, 8, 9, 11, 12, 12, 13, 14, 15, 16, 17, 20, 21, 22, 23, 24, 25, 27, 29, 32, 34, 35, 36, 39, 40, 41, 42, 44, 45, 47, 48, 49, 50.
- Butir soal dengan perolehan IDB kurang dari 0,25 hingga mencapai minus. Yakni butir soal yang harus diganti karena menyalahi logika, bahwa kelompok rendah lebih banyak menjawab benar daripada kelompok tinggi. Berikut nomor butir soal nya :  
4, 5, 7, 9, 14, 15, 17, 23, 24, 27, 36, 40, 50.
- Butir soal yang masih bisa di revisi karena perolehan ITK dan IDB hampir mencapai kriteria adalah nomor :  
6, 8, 10, 12, 13, 16, 20, 22, 32, 34, 35, 39, 42, 44, 45, 50.

## PENUTUP SIMPULAN

Dari hasil penelitian yang telah dilakukan di SMAN 12 dengan judul Analisis Butir Soal Ujian akhir Semester Ganjil maka hasil yang didapat adalah , telah diketahui bahwa dari 50 butir soal multiple choice Ujian Akhir Semester Ganjil bahasa Jerman SMAN 12 Surabaya terdapat 21 butir soal telah memenuhi syarat ITK, dan 13 butir soal telah memenuhi syarat IDB. Dan hasil akhir dari perhitungan indeks tingkat kesulitan dan indeks daya beda diketahui hanya 12 butir soal yang dapat dinyatakan layak.

Dari hasil analisis tingkat kesulitan butir soal, telah terbagi 3 golongan soal yaitu mudah, sedang dan sulit. Dari soal Ujian Akhir Semester Ganjil ini diketahui 3 soal tergolong sulit, 7 soal tergolong sedang dan 40 soal tergolong mudah. Namun dari 40 soal tergolong mudah ini telah di analisis bahwa 29 soal tergolong terlalu mudah sehingga dinyatakan tidak berfungsi. Lalu untuk hasil analisis indeks daya beda telah diketahui bahwa 37 butir soal tidak layak, dan diketahui juga dari soal tidak layak tersebut terdapat 13 butir soal yang indeks daya beda yang minus yaitu kelompok rendah lebih banyak menjawab benar daripada kelompok tinggi sehingga 13 soal tersebut tidak bisa di revisi dan diharapkan untuk dibuang dan diganti soal yang layak. Untuk beberapa sisa soal yang lain bisa untuk di revisi atau diperbaiki sehingga layak.

Sedangkan untuk analisis distraktor yang disediakan di 50 butir soal ini terdapat 250 opsi-opsi jawaban dan pengecoh, dan diketahui ada 133 opsi jawaban yang berfungsi baik , dan sisa 117 tidak berfungsi dengan baik karena tidak ada peserta ujian yang memilih satupun opsi jawaban tersebut. Untuk opsi jawaban yang kurang berfungsi baik, bisa direvisi atau diperbaiki.

## SARAN

Suatu butir soal tidak dapat dianggap valid dan tidaknya hanya melalui seberapa peserta didik banyak menjawab, tetapi peran pendidik dalam menyampikan materi dan membuat soal sangatlah penting. Suatu butir soal yang mudah dan hampir seluruh peserta

uji nya mampu mengerjakan juga dinamakan tidak valid, karena soal tersebut dianggap terlalu mudah bagi peserta uji. Namun jika soal tersebut juga tidak ada yang bisa mengerjakan atau terlalu sulit bagi peserta uji, soal tersebut juga dianggap tidak valid. Jadi soal yang valid atau layak adalah suatu soal yang memiliki Indeks tingkat kesulitan (ITK) dan indeks daya beda (IDB) yang memenuhi syarat standart atau kriteria. Butir soal yang baik adalah butir soal yang mampu mengukur dan memberikan gambaran sejauh mana kemampuan peserta didik. Maka dari itu sebelum melakukan tes atau ujian bagi peserta didik, diharapkan memperhatikan apakah butir-butir soal tersebut sudah layak dan mampu memenuhi kriteria Indeks Tingkat Kesulitan (ITK) dan Indeks Daya Beda (IDB) , maupun distraktor pengecohnya. Sehingga butir soal tersebut dapat berfungsi dengan baik untuk mengukur sejauh mana pemahaman dan kemampuan setiap peserta didik.

## DAFTAR PUSTAKA

- Arifin, Zainal. 2008. *Metodologi Penelitian Pendidikan*. Surabaya : Lentera Cendikia.
- Arikunto, Suharsimi Drs. 2005. *Dasar-dasar Evaluasi Pendidikan*. Jakarta: PT Bumi Aksara
- Daryanto. 2007. *Evaluasi Pendidikan*. Jakarta: PT Rineka Cipta.
- Nurgiyantoro Burhan. 2010. *Penilaian Dalam Pengajaran Bahasa Dan Sastra*, Yogyakarta : BPPE
- Samid Hasan. 2009. *Evaluasi Kurikulum* . Bandung: PT Remaja Rosdakarya.
- Sudijono, Anas. 2011. *Pengantar Evaluasi Pendidikan*. Jakarta: PT Rajagrafindo Persada.
- Subali, Bambang. 2012. *Prinsip Asesmen & Evaluasi Pembelajaran*. Yogyakarta: UNY Press.
- Surapranata, Sumarna. 2005. *Analisis, Validitas, Reliabilitas dan*

*Intrepretasi Hasil Tes.* Bandung: PT  
Remaja Rosdakarya.



# **ANALYSE VON ABSCHLUSSPÜFUNGSFRAGEN FÜR DEUTSCHE FÄCHER IN KLASSE 10 VON SMAN 12 SURABAYA**

**Elsa Wandira Maulidatirrahma**

Pädagogik der Deutsche Sprache, Fakultät für Sprache und Kunst, Universität Negeri  
Surabaya

[elsamaulidatirrahma16020094033@mhs.unesa.ac.id](mailto:elsamaulidatirrahma16020094033@mhs.unesa.ac.id)

**Drs. Suwarno Imam Samsul, M.Pd.**

Pädagogik der Deutsche Sprache, Fakultät für Sprache und Kunst, Universität Negeri  
Surabaya

## **AUSZUG**

Um zu messen, inwieweit die Fähigkeit und das Verständnis der Schüler für das Material, das während eines Semesters gegeben wurde, vorhanden ist, wird der Pädagoge am Ende des Semesters einen Test für Schüler geben, die dann anhand der Ergebnisse einen Überblick über den Pädagogen geben können wie weit ist der Erfolg von Lernzielen. Ein Test kann als gut oder machbar bezeichnet werden, wenn der Test auf Gültigkeit, Interpretation, Angemessenheit, Nützlichkeit, Vertrauenswürdigkeit und Wirksamkeit von Gegenständen mit einem guten Schwierigkeitsgrad und unterschiedlicher Leistung geprüft werden kann, da nicht alle Gegenstände diese Kriterien erfüllen können und kann gesagt werden, machbar zu sein, um Schüler zu geben. Die Formulierung des Problems in dieser Studie ist, wie der Schwierigkeitsgrad der Gegenstände, die unterschiedliche Kraft und Wirksamkeit des Antwort-Ablenkens von der Klasse X der Deutschen Abschlussprüfung (UAS) in SMAN 12 Surabaya TP 2019/2020. Die Forschung, die ich bei SMAN 12 Surabaya durchgeführt habe, ist eine quantitative Studie, bei der der Schwierigkeitsgrad, die unterschiedliche Leistung und die Wirksamkeit des Distraktors mithilfe der Formel IF (Index Facility) und ID (Index Discrimination) analysiert wurden. Aus den Forschungsergebnissen von SMAN 12 mit dem Titel Analyse der Prüfung zum Abschlussprüfungsgegenstand geht hervor, dass von den 50 Multiple-Choice-Fragen am Ende des Deutsch-Semesters bei SMAN 12 Surabaya 21 Punkte erfüllt sind. Die ITK-Anforderungen und 13 Elemente haben die IDB-Anforderungen erfüllt. Und die endgültigen Ergebnisse der Berechnung des Schwierigkeitsgradindex und des unterschiedlichen Leistungsindex sind nur 12 Elemente bekannt, die für machbar erklärt werden können. Aus den Ergebnissen der Analyse des Schwierigkeitsgrades der Gegenstände wurden 3 Gruppen von Fragen unterteilt, nämlich leicht, mittel und schwierig. Aus den Prüfungsfragen des letzten Semesters ist bekannt, dass 3 Fragen als schwierig, 7 Fragen als mittel und 40 Fragen als einfach eingestuft werden. Von den 40 als einfach eingestuften Fragen wurde jedoch analysiert, dass 29 Fragen als zu einfach eingestuft wurden, um für nicht funktionsfähig erklärt zu werden. Für die Ergebnisse der Analyse des unterschiedlichen Leistungsindex ist bekannt, dass 37 Elemente nicht durchführbar sind, und aus den falschen Fragen ist auch bekannt, dass es 13 Elemente mit einem Minus-Leistungsindexunterschied gibt,

nämlich dass die Antworten der niedrigen Gruppe korrekter sind als die hohe Gruppe, so dass 13 Fragen nicht überarbeitet werden können und voraussichtlich verworfen und durch entsprechende Fragen ersetzt werden. Für einige der verbleibenden Fragen kann es überarbeitet oder korrigiert werden, so dass es machbar ist. Während für die Analyse der in diesen 50 Elementen bereitgestellten Distraktoren 250 Antwort- und Täuschungsoptionen und 133 Antwortoptionen vorhanden waren, die gut funktionierten, funktionierten die verbleibenden 117 nicht richtig, da keiner der Testteilnehmer eine der Antworten auswählte Optionen. Für Antwortoptionen, die nicht gut funktionieren, können sie überarbeitet oder korrigiert werden. Schlüsselwörter:

*Schlüsselwörter: Problemanalyse, Schwierigkeitsgrad, Unterscheidungsmerkmale und Ablenker*

### **ABSTRACT**

To measure the extent to which the ability and understanding of students of the material that has been given during 1 semester, then at the end of the semester the educator will give a test to students who then from the results can provide an overview of the educator how far the success of learning objectives. A test can be said to be good or feasible if the test can be accounted for validity, interpretation, appropriateness, usefulness, trustworthiness, and effectiveness of items that include a good level of difficulty and different power, because not all items can be in accordance with these criteria. The formulation of the problem in this research is how the level of difficulty of the items, the different power and effectiveness of the answer distractor from the German Final Examination (UAS) class X in SMAN 12 Surabaya TP 2019/2020. The research that I did at SMAN 12 Surabaya is a quantitative study by analyzing the level of difficulty, different power and effectiveness of the distractor by using the formula IF (Index Facility) and ID (Index Discrimination). From the results of the research that has been carried out, it is known that of the 50 multiple choice questions at the end of the German language semester examination at SMAN 12 Surabaya there are 21 items that have met ITK requirements, and 13 items have met IDB requirements. And the final results of the calculation of the difficulty level index and the different power index are known only 12 items that can be declared feasible. From the results of the analysis of the difficulty level of the items, 3 groups of questions were divided namely easy, medium and difficult. From this UAS, 3 problems are classified as difficult, 7 questions are classified as moderate and 40 questions are classified as easy. However, of the 40 questions classified as easy, it has been analyzed that 29 questions were classified as too easy to declare not functioning. Then for the results of the analysis of the different power index it is known that 37 items are not feasible, and it is also known from the improper questions that there are 13 items with a minus power index difference, namely the low group answers more correctly than the high group so that 13 questions cannot be revised and expected to be discarded and replaced with appropriate questions. For some of the remaining questions, it can be revised or corrected so that it is feasible. Whereas for the analysis of the distractors provided in these 50 items, there were 250 answer and deception options, and there were 133 answer options that functioned well, and the remaining 117 did not function properly because none of the test participants chose any of the answer options. For answer options that don't work well, you can fix them.

*Keywords: item analysis, level of difficulty, different power and distractor*

## VORLÄUFIG

Um das Niveau der Fähigkeiten und des Verständnisses der Schüler zu messen, bieten die Pädagogen den Schülern Tests an. Anschließend können die Ergebnisse dieser Tests diagnostiziert werden, wenn der Ort der Lernschwierigkeiten der Schüler liegt und die Wissensfähigkeiten der Schüler bekannt sind, und Informationen über den Erfolg bereitstellen Lernziele der Schüler zu erreichen. Aber um den Test messen zu können, muss gesagt werden, dass er machbar oder gut ist. Die Merkmale eines guten Tests müssen ausgehend von Gültigkeit (Gültigkeit), Durchführbarkeit, Interpretation, Vertrauenswürdigkeit (Zuverlässigkeit), Nützlichkeit und Wirksamkeit von Elementen einschließlich des Schwierigkeitsgrades der Fragen und der Fähigkeit zur Unterscheidung guter Fragen berücksichtigt werden.

SMAN 12 Surabaya bietet seinen Schülern den UAS-Test (Final Semester Examination) an, um herauszufinden und zu messen, inwieweit SMAN 12 Surabaya Schüler das Material in einem Semester beherrschen können. Daher muss auch die Qualität der Fragen geprüft werden, indem der Schwierigkeitsgrad, die Unterscheidungsmerkmale und die Funktion des zu testenden Gegenstands des Gegenstands getestet werden. Ein Semester. Die Formulierung des Problems in dieser Studie lautet: Wie ist der Schwierigkeitsindex, der Differenzleistungsindex und die Wirksamkeit von Distraktoren für jedes Element? Diese Studie zielt darauf ab, den Schwierigkeitsindex, den Differenzleistungsindex und die Wirksamkeit von Distraktoren für jeden Punkt im Abschlussprüfungssemester 12 SMA Negeri Surabaya zu beschreiben. Die Einschränkung des Problems in dieser Studie besteht darin, dass sich diese Studie darauf konzentriert, den Schwierigkeitsindex, den Differenzleistungsindex und die Wirksamkeit von Distraktoren für jedes Multiple-Choice-Element mithilfe einer quantitativen Analyse

zu ermitteln. Der Gegenstand dieser Forschung konzentriert sich auf die Fragen der deutschen XAS-Klasse X SMAN 12 Surabaya.

## METHODE

Diese Art von Forschung, die bei SMAN 12 Surabaya durchgeführt wird, ist eine Art quantitativer Forschung. Wie bereits erwähnt, handelt es sich bei der quantitativen Forschung von Arikunto (2002: 10) um eine Studie, deren Messergebnisse erforderlich sind, um viele Zahlen zu verwenden, z. B. den Prozess der Datenerfassung für das Erscheinungsbild der Ergebnisse. Die Forschungsmethode wurde in einer quantitativen Analyse mit Forschungsverfahren durchgeführt, nämlich:

1. Vorbereitungsphase  
Verwendung von Literaturstudien, nämlich Auffinden und Abstumpfen von Theorien, die für das zu untersuchende Thema relevant sind, und Erteilen von Datenerfassungsgenehmigungen.
2. Phase der Datenerfassung  
Die Daten wurden im Schuljahr SMAN 12 Surabaya Klasse X 2019/2020 aufgenommen. Die Daten liegen in Form von Fragebögen und Antwortschlüsseln vor.
3. Datenverarbeitungsphase  
Die gesammelten Daten werden dann unter Verwendung quantitativer statistischer Analysetechniken analysiert und dann zur Überarbeitung beschrieben.
4. Berichtsvorbereitungsphase  
Nachdem alle Daten analysiert wurden, wird ein Bericht erstellt, um die Forschungsziele zu beantworten

Die quantitative Analyse in dieser Studie dient der Analyse des Schwierigkeitsgrades und der Unterscheidung von Problemen mit der Formel.

Die Schritte zur Analyse der Elemente gemäß Nurgiyantoro (2001: 136) sind wie folgt:



A. Sortieren Sie die Punktzahlen auf dem Antwortbogen der Schüler von der höchsten Punktzahl in einer Reihe nach unten.

B. Nachdem die Daten sortiert wurden, werden sie in zwei Gruppen unterteilt: hohe und niedrige Gruppen. Zum Beispiel beträgt die Anzahl der Schüler 40 Personen, dann sind die 20 Schüler mit der höchsten Punktzahl die hohe Gruppe und die 20 Schüler mit der niedrigsten Punktzahl die niedrige Gruppe.

C. Nachdem zwei Gruppen gebildet wurden, können die hohe und die niedrige Gruppe den ITK-Punkt (Difficulty Index) berechnen, der anhand der Schülerantworttabelle mit der folgenden Formel erstellt werden kann:

$$ITK = \frac{FKT + FKR}{N}$$

Beschreibung :

ITK = Index des gesuchten Schwierigkeitsgrades

FKT = Hohe Anzahl korrekter Antworten in Gruppen

FKR = Die Anzahl der richtigen Antworten von Gruppen ist gering

N = Anzahl der Testteilnehmer für beide Gruppen.

Nach dem Ausfüllen des Schwierigkeitsindex (ITK) können Sie den Differenzleistungsindex (IDB) mit einer Formel wie der folgenden berechnen:

$$IDB = \frac{FKT - FKR}{N}$$

Beschreibung :

IDB = Differenzleistungsindex gesucht

FKT = Hohe Anzahl korrekter Antworten in Gruppen

FKR = Die Anzahl der richtigen Antworten von Gruppen ist gering

N = Anzahl der Testteilnehmer für beide Gruppen.

Nachdem Sie die Elemente für den Schwierigkeitsgradindex (ITK) und den Differenzleistungsindex (IDB) ermittelt haben, müssen Sie im nächsten Schritt nach der Effektivität des Distraktors suchen. Dazu wird eine Tabelle mit der Verteilung der

Häufigkeit der Antworten von Schülern in hohen und niedrigen Klassen erstellt Gruppen. Hier ist eine Beispieltabelle:

NO	Hohe gruppe			
	A	B	C	D
1	*10	2	2	2
2	5	-	-	*15
3	6	*7	3	1
4	7	*6	6	2
5	9	-	*7	5
Dst...				

NO	niedrige Gruppe			
	A	B	C	D
1	*10	4	6	2
2	8	*7	3	3
3	7	2	4	*6
4	6	*8	2	4
5	3	5	*7	5
Dst...				

Hinweis:

- A, B, C, D = Antwortoptionen

- (\*) = richtige Antwort

- Oben sind Beispiele von 20 Schülern hoher Gruppen und 20 Schülern niedriger Gruppen aufgeführt.

Nachdem die Schüler die Auswahl der Antwortverteilungsdaten erhalten haben, können wir eine Distraktoranalyse für jeden Gegenstand durchführen, um die Wirksamkeit des Distraktors oder der Täuschung zu bestimmen. Laut Nurgiyantoro (2010: 2010) gibt es mehrere wirksame Ablenkungskriterien, nämlich:

1. Alle Distraktoren (falsche Option) müssen wählen
2. Die Anzahl der Wähler mit falscher Wahl aus Teilnehmern hoher Gruppen muss geringer sein als die der niedrigen Gruppe
3. Wenn es nur einen Wähler mit falscher Wahl gibt, muss er aus einer niedrigen Gruppe stammen
4. Die Kriterien 2 und 3 werden manchmal als belastend angesehen, die sich im Wesentlichen nicht von der Logik der oben genannten IDB-Anforderungen unterscheiden. Was



also effektiv verwendet wird, sind die Kriterien 1.

Weniger effektive Distraktorooptionen müssen verworfen oder überarbeitet werden. Effektive Distraktorooptionen sind Optionen, die unauffällig sind und der ursprünglichen Antwort fast ähnlich sind, oder "ähnlich, aber nicht gleich", sodass weniger vorsichtige Lernende die Distraktorooption wählen.

## DISKUSSION

Die Artikelanalyse hat das Ziel, jede Frage anhand von guten, fehlenden oder schlechten Fragen zu identifizieren. Um die Qualität der Fragen herauszufinden, müssen der Schwierigkeitsindex (ITK) und der Differenzleistungsindex (IDB) bekannt sein. Um die ITK- und IDB-Elemente herauszufinden, müssen Sie in mehreren Schritten vorgehen:

1. Der erste Schritt besteht darin, eine Tabelle zu erstellen, die die Initialen der Fächer der Schüler enthält und die Werte und Antworten enthält, die aus den Korrekturergebnissen, die jeder Schüler als Ganzes erhalten hat, wahr oder falsch sind, indem eine Punktzahl von 1 für die gegebene Richtige Antwort und Punktzahl 0 für die richtige Antwort ist falsch.
2. Der zweite Schritt besteht darin, eine Tabelle zu erstellen, die die Initialen der Fächer der Schüler enthält und dann von den Schülern mit der höchsten Anzahl korrekter Antworten oder der geringsten Anzahl von Fehlern zu den Schülern mit den wenigsten richtigen Antworten oder Antworten sortiert wird die höchste Anzahl von Fehlern.
3. Nachdem die Datenergebnisse der Noten der Schüler sortiert wurden, beginnend mit den meisten, haben sie die richtigsten Antworten oder haben die wenigen Fehler zu den richtigen Antworten, die die wenigsten oder mehr falschen Antworten haben, und können danach in zwei Gruppen unterteilt werden. Nämlich die hohe und die niedrige Gruppe. Die hohe Gruppe ist die Hälfte der Gesamtzahl der

Schüler mit der höchsten Punktzahl, während die niedrige Gruppe der Rest der oberen Hälfte ist. Die Anzahl der Schüler in einer Klasse besteht aus 36 Schülern, daher ist die Reihenfolge der Schüler Nr. 1-18 oben die hohe Gruppe, während die Schülerzahlen 19-36 niedrige Gruppen sind.

4. Nachdem Sie hohe und niedrige Gruppen gebildet haben, müssen Sie als Nächstes den Schwierigkeitsindex (ITK) kennen, indem Sie ihn mithilfe einer Formel berechnen.
5. Nachdem Sie den Schwierigkeitsgradindex der Elemente berechnet und kennen, können Sie ihn fortsetzen, indem Sie die IDB-Elemente (Difference Power Index) mithilfe der Formel berechnen.
6. Nach der Berechnung des Schwierigkeitsgradindex (ITK) und des Differenzleistungsindex (IDB) wird als Nächstes eine kombinierte Tabelle der Ergebnisse der Berechnung von ITK und IDB erstellt. Geeignete Fragen sind Fragen, die die Anforderungen von ITK und IDB erfüllen können. Wenn sie nur die Anforderungen von ITK erfüllen und die IDB-Anforderungen nicht erfüllen können, wird das Problem als unangemessen bezeichnet und umgekehrt. Um die ITK-Anforderungen erfüllen zu können, muss jeder Artikel eine ITK von 0,25 bis 0,85 aufweisen. Je höher die ITK, desto einfacher das Problem, desto niedriger die ITK, desto schwieriger ist das Problem. Um die IDB-Anforderungen erfüllen zu können, muss jedes Element eine IDB über 0,25 haben. Je höher die IDB, desto besser funktioniert die Funktion, da hohe und niedrige Gruppen unterschieden werden können. Bei Fragen, die die ITK- oder IDB-Anforderungen nicht erfüllen, können sie überarbeitet oder korrigiert werden.

## ERGEBNISSE

Jeder Artikel wird für würdig erklärt oder kann nicht aus ITK (Schwierigkeitsindex) und IDB (Differenzleistungsindex) gesehen werden, die die angegebenen Anforderungen erfüllen oder nicht erfüllen. Wenn ein Gegenstand die Anforderung des Schwierigkeitsgradindex erfüllt hat, aber die Anforderung des unterschiedlichen Leistungsindex nicht erfüllen kann, gilt der Gegenstand als fehlerhaft oder nicht ordnungsgemäß. Im Gegenteil, wenn die Elemente die Anforderungen für den Differenzleistungsindex erfüllen können, aber die Schwierigkeitsindexanforderung nicht erfüllen können, gelten die Elemente als nicht funktionsfähig oder unangemessen. Jeder Artikel, der als angemessen und gut angesehen wird, verfügt über ITK und IDB, die die Anforderungen erfüllen. In Verbindung damit, den Ergebnissen der Analyse, die durchgeführt wurde, um den Schwierigkeitsindex, den Differenzleistungsindex und die Distraktoranalyse zu finden, wurden mehrere Ergebnisse gefunden, nämlich es gibt einige Punkte, von denen gesagt werden kann, dass sie machbar oder funktionsfähig sind, und es gibt auch Elemente, von denen gesagt werden kann, dass sie nicht richtig funktionieren. Als nächstes können die Fragen, die als unangemessen und nicht funktionsfähig gelten, oder die Elemente, deren ITK und IDB noch nicht berechtigt sind, überarbeitet oder ersetzt werden. Einige Elemente, die die ITK- und IDB-Anforderungen erfüllen, müssen nicht überarbeitet oder ersetzt werden. In Übereinstimmung mit den Ergebnissen der durchgeführten Analyse ist aus insgesamt 50 Elementen bekannt, dass es 29 Elemente gibt, die die Anforderungen des Schwierigkeitsindex nicht erfüllen. Dann ist auch bekannt, dass 37 Elemente die Anforderungen des Schwierigkeitsindex nicht erfüllen Anforderungen an den Differenzleistungsindex. Die Endergebnisse sind alle bekannt, dass es 12 Elemente gibt, die die Anforderungen des ITK (Schwierigkeitsindex) und des IDB (Differenzleistungsindex) erfüllen. Dann

können die 12 Elemente als machbar bezeichnet werden, und die verbleibenden 38 Elemente werden als nicht durchführbar angesehen Da sie die ITK- und IDB-Anforderungen nicht erfüllen können, können die 38 Elemente überarbeitet oder durch bessere Fragen ersetzt werden. In den SMAN 12 Surabaya-Abschlusssemesterprüfungsfragen sind 29 Elemente enthalten, die nicht dem Schwierigkeitsindex entsprechen. Daher kann der Schluss gezogen werden, dass die Abschlusssemesterprüfungsfragen für Schüler immer noch zu einfach sind. Dann gibt es 37 Punkte, die die unterschiedlichen Power Index-Anforderungen nicht erfüllen können. Es kann gefolgert werden, dass die Prüfungsfragen des letzten Semesters nicht gut funktionieren, um zwischen Schüler mit hoher und niedriger Gruppe unterscheiden zu können.

Es gibt mehrere Punkte, bei denen die Ergebnisse der ITK- und IDB-Analyse die Kriterienanforderungen fast erreicht haben. Diese Punkte können dann noch erneut überarbeitet werden. Es gibt aber auch Elemente, die wirklich ersetzt und sogar verworfen werden müssen, da die Übernahme von ITK und IDB alles andere als machbar ist. Hier sind einige Elemente, die überarbeitet oder verworfen werden müssen:

- Elemente, die mit dem Erwerb von ITK 0,85-1,00 zu einfach sind:  
1, 3, 4, 5, 6, 9, 11, 12, 13, 14, 15, 16, 17, 21, 22, 23, 24, 27, 34, 35, 36, 39, 40, 41, 42, 44, 47, 49, 50.
- Artikel mit dem Erwerb von IDB unter 0,25:  
1, 3, 4, 5, 6, 7, 8, 9, 11, 12, 12, 13, 14, 15, 16, 17, 20, 21, 22, 23, 24, 25, 27, 29, 32, 34, 35, 36, 39, 40, 41, 42, 44, 45, 47, 48, 49, 50.
- Artikel mit IDB-Erfassung ist kleiner als 0,25, um Minus zu erreichen. Nämlich die Elemente, die ersetzt werden müssen, weil sie gegen die Logik verstoßen, dass die niedrige Gruppe korrekter antwortet als die

hohe Gruppe. Hier sind die Artikelnummern:  
4, 5, 7, 9, 14, 15, 17, 23, 24, 27, 36, 40, 50.

## **VORSCHLAGEN**

Ein Artikel kann nicht als gültig angesehen werden und nicht nur durch die Anzahl der Schüler, sondern auch durch die Rolle der Pädagogen bei der Bereitstellung von Material und bei der Beantwortung von Fragen. Ein Element, das einfach ist und fast alle Testteilnehmer arbeiten können, wird auch als ungültig bezeichnet, da die Frage für Testteilnehmer als zu einfach angesehen wird. Wenn die Frage jedoch auch von niemandem gestellt werden kann oder für Testteilnehmer zu schwierig ist, wird die Frage ebenfalls als ungültig angesehen. Eine gültige oder realisierbare Frage ist also eine Frage mit einem Schwierigkeitsgradindex (ITK) und einem anderen Leistungsindex (IDB), der die Standardanforderungen oder -kriterien erfüllt. Gute Gegenstände sind Gegenstände, die messen und eine Vorstellung davon geben können, inwieweit die Fähigkeiten der Schüler sind. Vor der Durchführung von Tests oder Prüfungen für Schüler wird daher erwartet, dass darauf geachtet wird, ob die Elemente geeignet sind und die Kriterien des Schwierigkeitsindex (ITK) und des Differenzleistungsindex (IDB) sowie der Ablenker erfüllen. Damit diese Elemente gut funktionieren können, um das Ausmaß des Verständnisses und der Fähigkeiten jedes Schülers zu messen. • Elemente, die noch überarbeitet werden können, weil die Übernahme von ITK und IDB die Kriterien fast erreicht hat, sind Zahlen:  
6, 8, 10, 12, 13, 16, 20, 22, 32, 34, 35, 39, 42, 44, 45, 50.

## **SCHLUSS ZUSAMMENFASSUNG**

Aus den Forschungsergebnissen von SMAN 12 mit dem Titel Analyse der Prüfung zum Abschlussprüfungsgegenstand geht hervor, dass von den 50 Multiple-Choice-Fragen am

Ende des Deutsch-Semesters bei SMAN 12 Surabaya 21 Punkte erfüllt sind. Die ITK-Anforderungen und 13 Elemente haben die IDB-Anforderungen erfüllt. Und die endgültigen Ergebnisse der Berechnung des Schwierigkeitsgradindex und des unterschiedlichen Leistungsindex sind nur 12 Elemente bekannt, die für machbar erklärt werden können.

Aus den Ergebnissen der Analyse des Schwierigkeitsgrades der Gegenstände wurden 3 Gruppen von Fragen unterteilt, nämlich leicht, mittel und schwierig. Aus den Prüfungsfragen des letzten Semesters ist bekannt, dass 3 Fragen als schwierig, 7 Fragen als mittel und 40 Fragen als einfach eingestuft werden. Von den 40 als einfach eingestuften Fragen wurde jedoch analysiert, dass 29 Fragen als zu einfach eingestuft wurden, um für nicht funktionsfähig erklärt zu werden. Für die Ergebnisse der Analyse des unterschiedlichen Leistungsindex ist bekannt, dass 37 Elemente nicht durchführbar sind, und aus den falschen Fragen ist auch bekannt, dass es 13 Elemente mit einem Minus-Leistungsindexunterschied gibt, nämlich dass die Antworten der niedrigen Gruppe korrekter sind als die hohe Gruppe, so dass 13 Fragen nicht überarbeitet werden können und voraussichtlich verworfen und durch entsprechende Fragen ersetzt werden. Für einige der verbleibenden Fragen kann es überarbeitet oder korrigiert werden, so dass es machbar ist.

Während für die Analyse der in diesen 50 Elementen bereitgestellten Distraktoren 250 Antwort- und Täuschungsoptionen und 133 Antwortoptionen vorhanden waren, die gut funktionierten, funktionierten die verbleibenden 117 nicht richtig, da keiner der Testteilnehmer eine der Antworten auswählte. Optionen. Für Antwortoptionen, die nicht gut funktionieren, können sie überarbeitet oder korrigiert werden.

## **BIBLIOGRAPHIE**



- Arifin, Zainal. 2008. *Metodologi Penelitian Pendidikan*. Surabaya : Lentera Cendikia.
- Arikunto, Suharsimi Drs. 2005. *Dasar-dasar Evaluasi Pendidikan*. Jakarta: PT Bumi Aksara
- Daryanto. 2007. *Evaluasi Pendidikan*. Jakarta: PT Rineka Cipta.
- Nurgiyantoro Burhan. 2010. *Penilaian Dalam Pengajaran Bahasa Dan Sastra*, Yogyakarta : BPPE
- Samid Hasan. 2009. *Evaluasi Kurikulum* . Bandung: PT Remaja Rosdakarya.
- Sudijono, Anas. 2011. *Pengantar Evaluasi Pendidikan*. Jakarta: PT Rajagrafindo Persada.
- Subali, Bambang. 2012. *Prinsip Asesmen & Evaluasi Pembelajaran*. Yogyakarta: UNY Press.
- Surapranata, Sumarna. 2005. *Analisis, Validitas, Reliabilitas dan Intrepretasi Hasil Tes*. Bandung: PT Remaja Rosdakarya.

